

new U.S. Appln. based on Appln. Nos. 60/236,574 & 60/299,360 for Online Intelligent Information... Express Mail Label No. EL864389877US Attorney Docket No.: 2102680-990101

Gray Cary et al. – GTS/415-836-2500 Applicants: Victor Hsieh Sheet 1 of 36 •

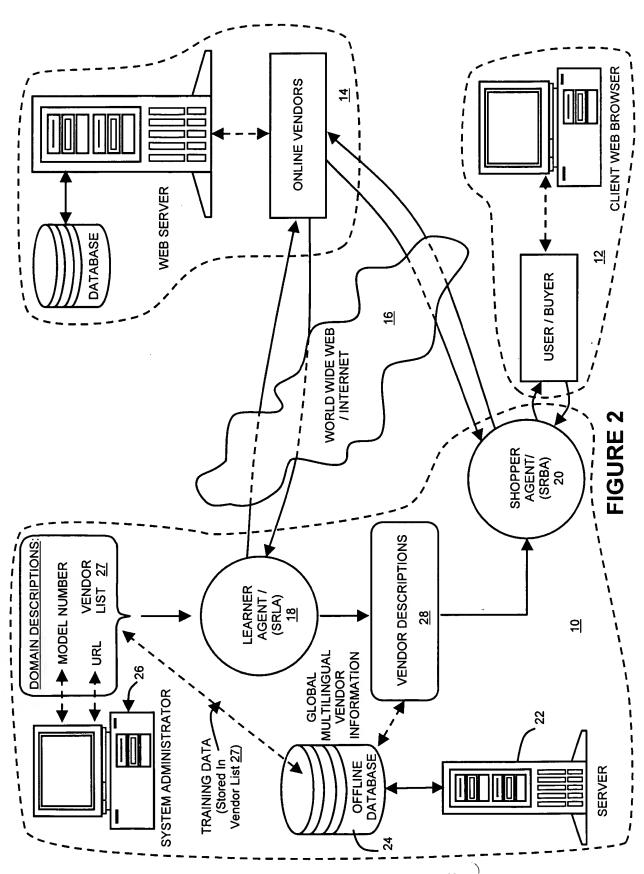
{7,","",""</I>,""</TABLE>"}

procedure pcwrapHLRT (page P)

scan to the line 7 in P

while the next occurrence of li in P occurs before the next occurrence of t scan in P to next occurrence of li; save position as start of item attribute scan in P to next occurrence of li; save position as end of item attribute scan in P to next occurrence of li; save position as start of price attribute scan in P to next occurrence of li; save position as end of price attribute return extracted , (item, price), } pairs

FIGURE 1 (PRIOR ART)







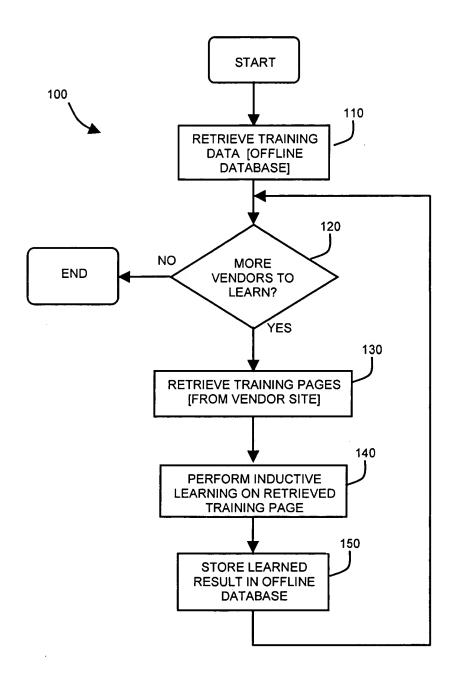


FIGURE 3



Element Name	Element Description
Vendor Name	Name of the online shop
Vendor URL	The URL of the online shop
Form URL	The URL to submit search data
Learning Domain	The domain used to learn the vendor description
Head	The end of header position of vendor's product pages. This element to learnt to reduce the product page searching space and thus shorten the product information extraction time
Tail	The start of footer position of vendor's product pages. Same as Head element, with this element the product information extraction time for Shopper Agent is shorten.
Left Delimiter of Item	The Shopper Agent uses this two delimiters to identify
Right Delimiter of Item	a product
Left Delimiter of Price	The Shopper Agent uses this two delimiters to locate
Right Delimiter of Price	a product's price information

FIGURE 4

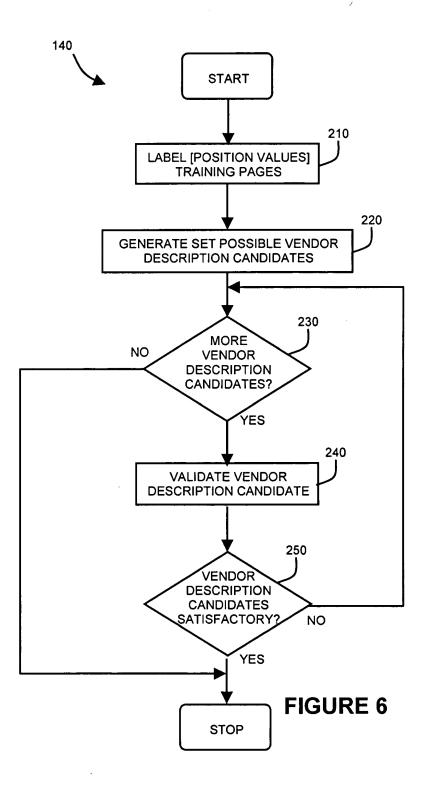
	EXAMPLE
Element	Element Description
Name	•
Vendor Name	800.com
Vendor URL	http://www.800.com
Form URL	http://www.800.com/search/srchrslts.asp?qs=1&slteentry=All&entry=
Learning	Md
Domain	
Head	1230
Tail	
Left Delimiter	/td width="12">
of Item	
Right	
Delimiter of	
Item	
Left Delimiter	Your Price: \$
of Price	
Right	
Delimiter of	
Price	

FIGURE 5



Applicants: Victor Hsieh

Sheet 5 of 36







MD PRICE		<html></html>		
		<title>A Simple Product Catalogs</title>		
Model	PRICE	<body></body>		
Number	(US\$)	<h2> MD Price </h2>		
HM381MD	399.95	<table border="1"> <tr bgcolor="ORANGE"><th> Model Number </th></tr></table>	Model Number	
Model Number				
MD2070	599.95	<tr bgcolor="ORANGE"><th> wioder Number </th></tr>	wioder Number	
wioder Number				
MD203	249.95	<tr><td> HM381MD</td><td><i>399.95</i></td></tr>	 HM381MD	<i>399.95</i>
 HM381MD	<i>399.95</i>			
MDR3	399.95	<tr><td> MD2070</td><td><i>599.95</i></td></tr>	 MD2070	<i>599.95</i>
 MD2070	<i>599.95</i>			
		<tr><td> MD203</td><td><i>249.95</i></td></tr>	 MD203	<i>249.95</i>
 MD203	<i>249.95</i>			
		<tr><td> MDR3</td><td><i>399.95</i></td></tr>	 MDR3	<i>399.95</i>
 MDR3	<i>399.95</i>			

		ADA For all Of The December 4 (December 1)		
		End Of The Product Catalog		
			VATIVILY	

FIGURE 7

Product Entry	Labels
{ HM381MD, 399.95}	{<<174, 180>, <197, 202>>}
{ MD2070, 599.95}	{<<229, 234>, <251, 256>>}
{ MD203, 249.95}	{<<283, 287>, <304, 309>>}
{ MDR3, 399.95}	{<<336, 339>, <356, 361>>}

FIGURE 8

Product Entry	Labels
{ PRODUCT, PRICE}	{ <product delimiter,<br="" left="">PRODUCT RIGHT DELIMITER>, <price delimiter,="" left="" price<="" td=""></price></product>
	RIGHT DELIMITER>}
* * *	* * *

FIGURE 9



Applicants: Victor Hsieh

Sheet 7 of 36

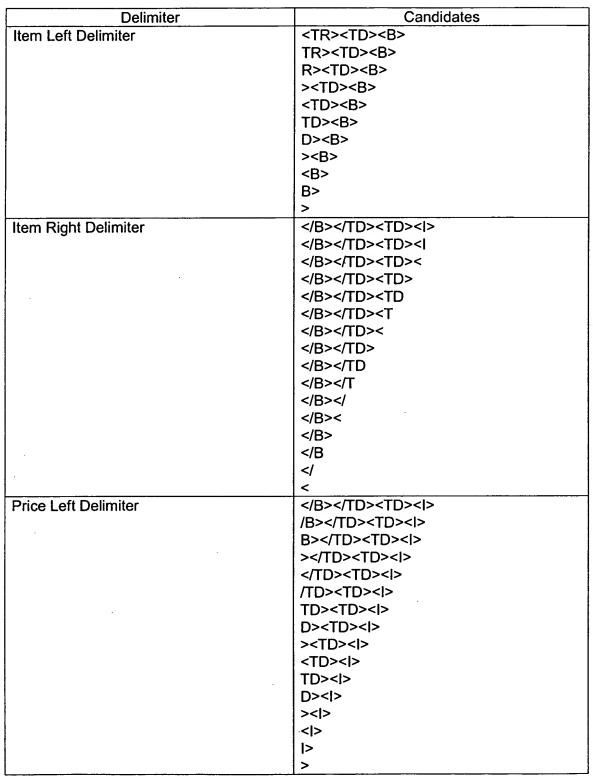


FIGURE 10A



Applicants: Victor Hsieh

Sheet 8 of 36

Delimiter	Candidates
Price Right Delimiter	
	/
	</td
	<
Head	<
	<h< td=""></h<>
	<ht< td=""></ht<>
	<htm< td=""></htm<>
	<html< td=""></html<>
	<html></html>
Tail	>
	L>
	ML>
	TML>
	HTML>
	/HTML>

FIGURE 10B

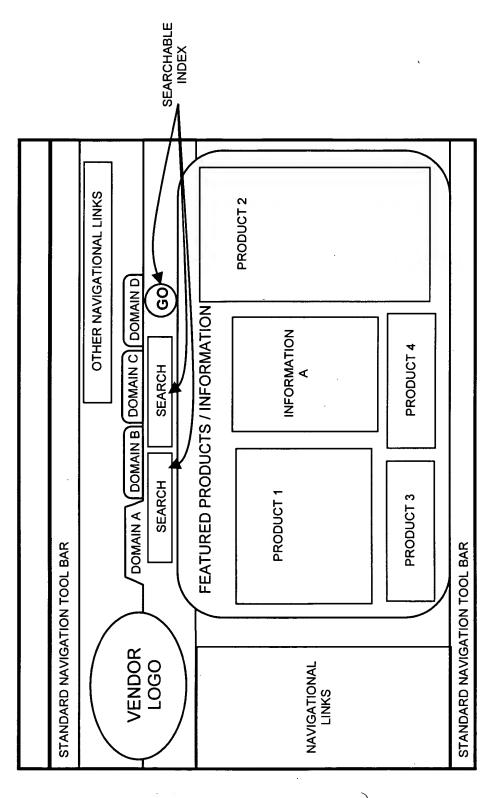


FIGURE 11

new U.S. Appln. based on Appln. Nos. 60/236,574 & 60/299,360 for Online Intelligent Information... Express Mail Label No. EL860-990101 Attorney Docket No.: 2102680-990101 Gray Cary et al. – GTS/415-836-2500 Gray Cary et al. – GTS/415-836-2500

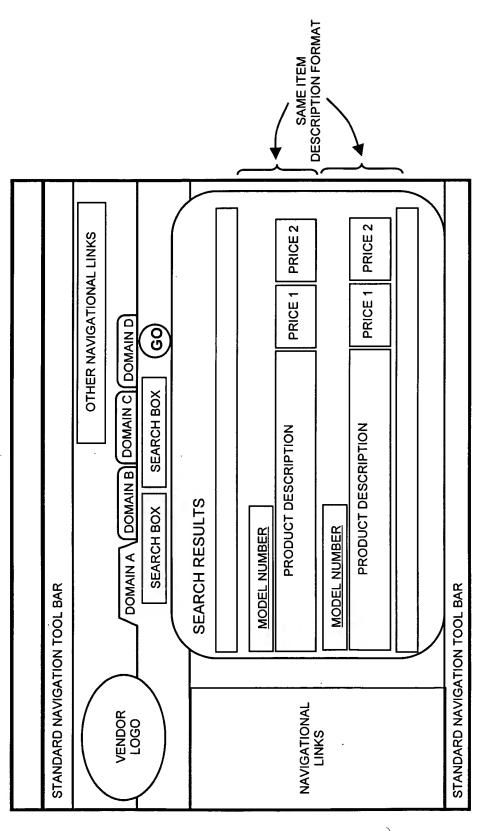


FIGURE 12

new U.S. Appln. based on Appln. Nos. 60/236,574 & 60/299,360 for Online Intelligent Information... Attorney Docket No.: 2102680-990101 Gray Cary et al. – GTS/415-836-2500 Gray Cary et al. – GTS/415-836-2500

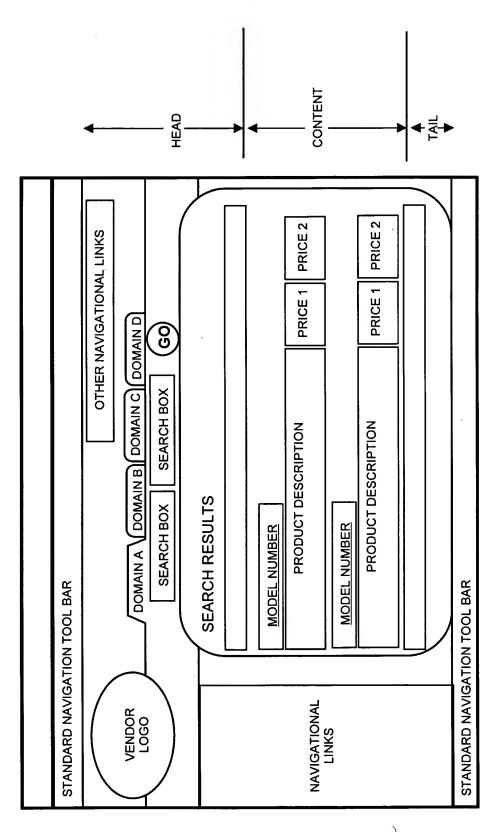
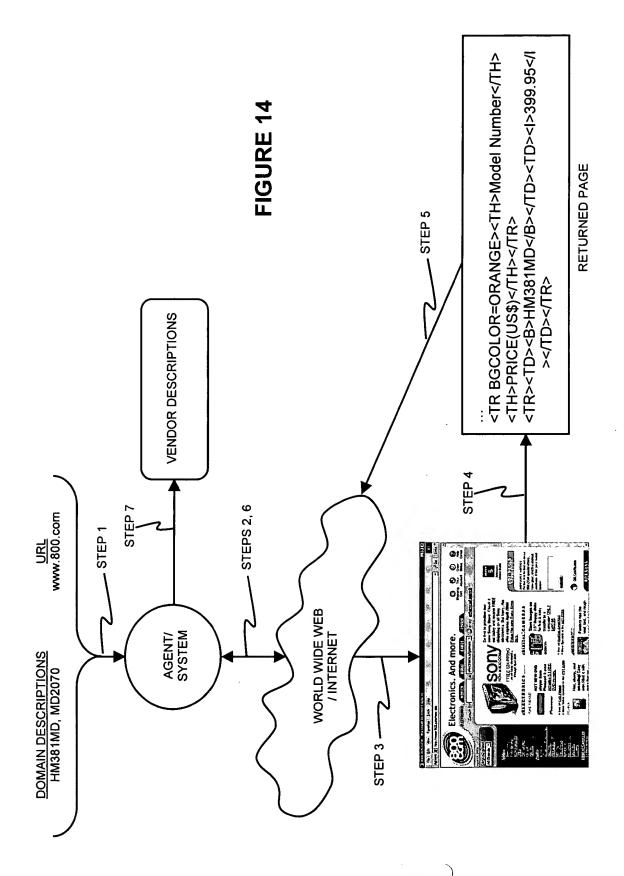


FIGURE 13

new U.S. Appln. based on Appln. Nos. 60/236,574 & 60/299,360 for Online Intelligent Information... Express Mail Label No.: 21026480-990101 Gray Cary et al. – GTS/415-836-2500 Applicants: Victor Hsieh Sheet II of 36





new U.S. Appln. based on Appln. Nos. 60/236,574 & 60/299,360 for Online Intelligent Information... Express Mail Label No. EL864389877US Attorney Docket No.: 2102680-990101

Gray Cary et al. – GTS/415-836-2500

Applicants: Victor Hsieh Sheet 13 of 36

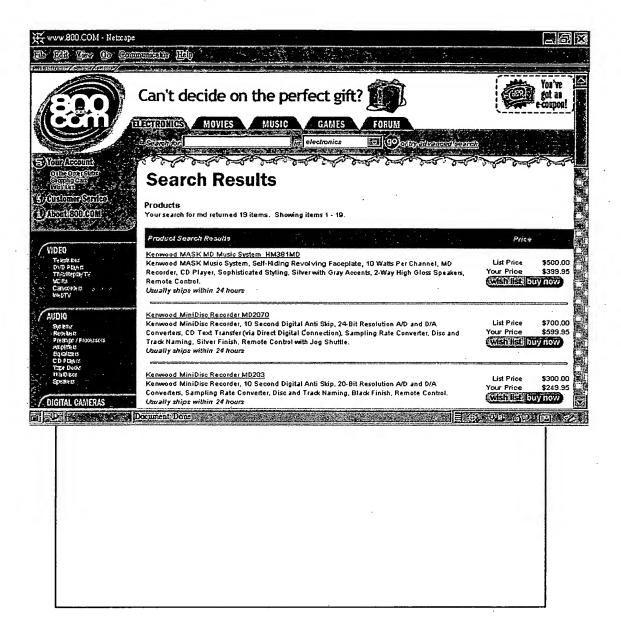
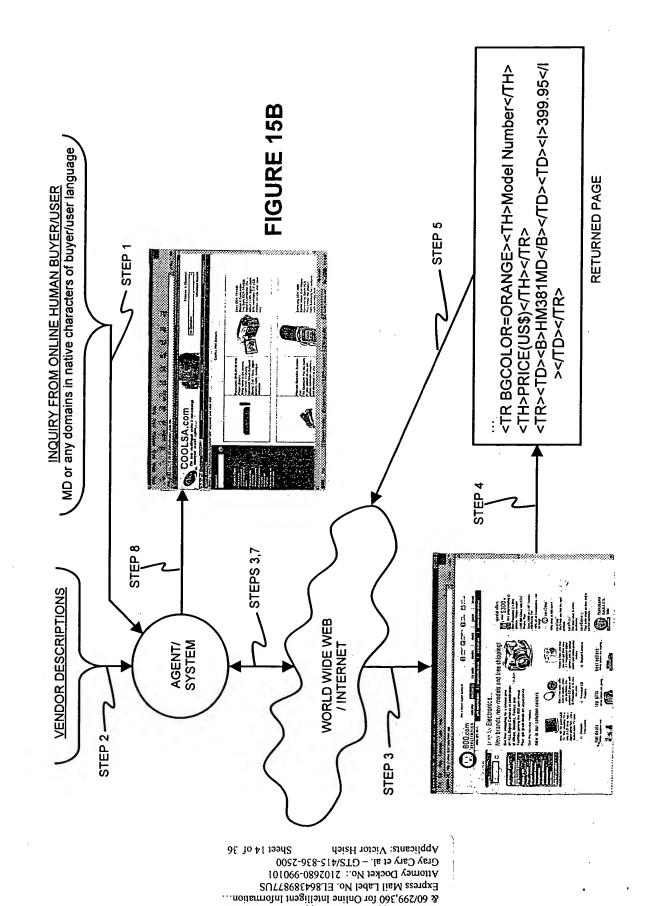
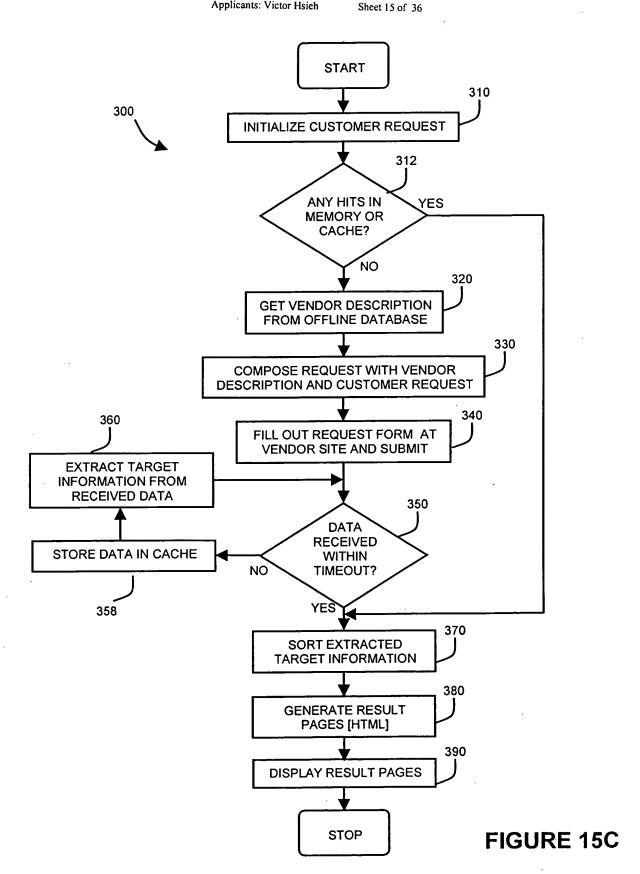


FIGURE 15A



new U.S. Appln. based on Appln. Nos. 60/236,574

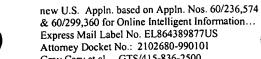




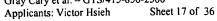


endor Information Add Vendor	
Section Patients Assessment of the Section Sec	
	Add Vendor
_ Details	
Vendor Name:	Provide Training Examples
URL:	
Form's URL	
Learning Domain:	
Head	Pat
Left Delimiter of Item:	Right Delimiter of Item:
Left Delimiter of Price:	Right Delimiter of Price:
	Save
	OK Cancel Apply

FIGURE 16



Gray Cary et al. - GTS/415-836-2500



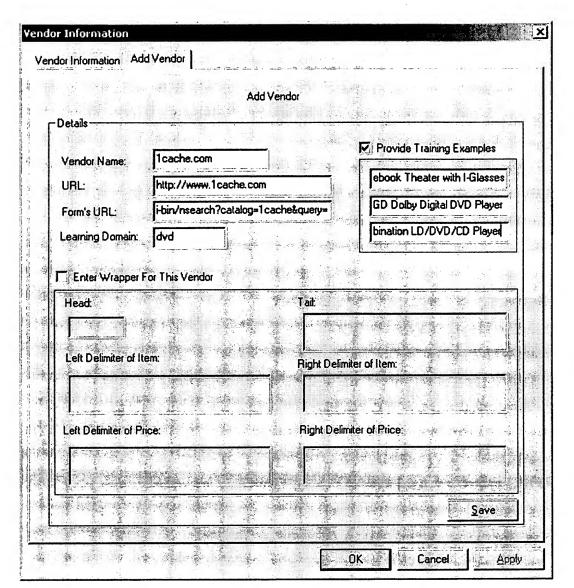


FIGURE 17



Applicants: Victor Hsieh Sheet 18 of 36

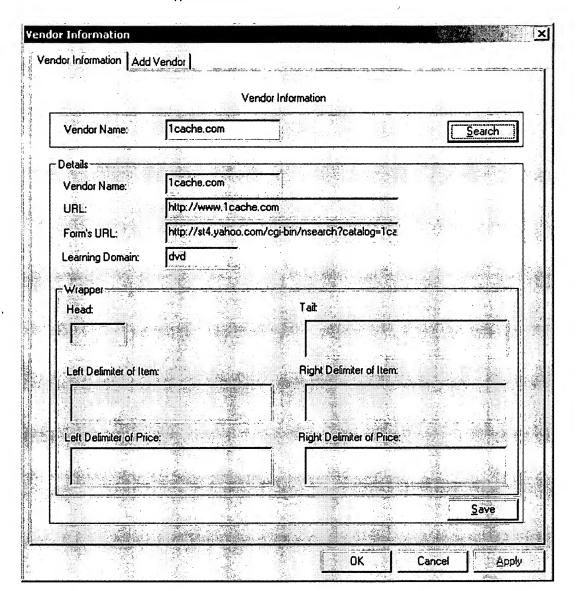
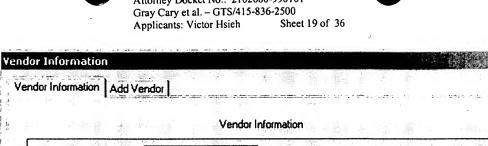


FIGURE 18



new U.S. Appln. based on Appln. Nos. 60/236,574 & 60/299,360 for Online Intelligent Information... Express Mail Label No. EL864389877US Attorney Docket No.: 2102680-990101



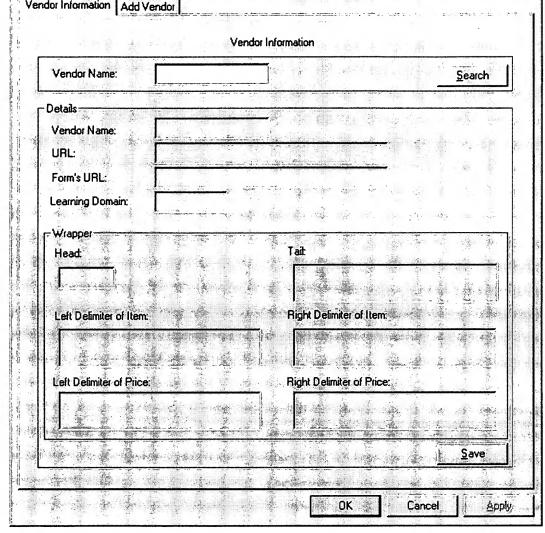


FIGURE 19



new U.S. Appln. based on Appln. Nos. 60/236,574 & 60/299,360 for Online Intelligent Information... Express Mail Label No. EL864389877US Attorney Docket No.: 2102680-990101

Gray Cary et al. – GTS/415-836-2500 Applicants: Victor Hsieh Sheet 20 of 36



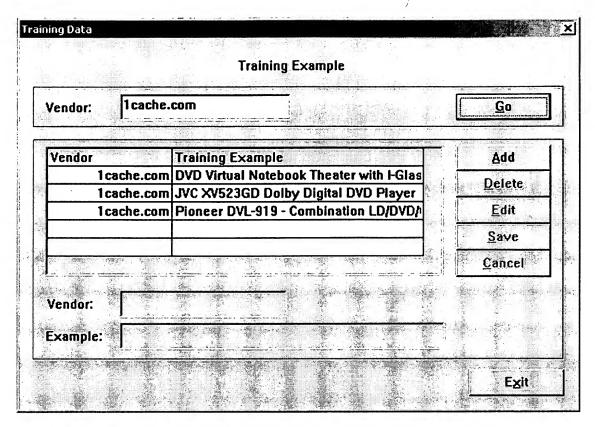


FIGURE 20

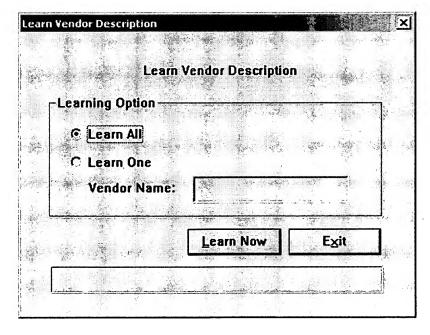


FIGURE 21



Applicants: Victor Hsieh Sheet 21 of 36



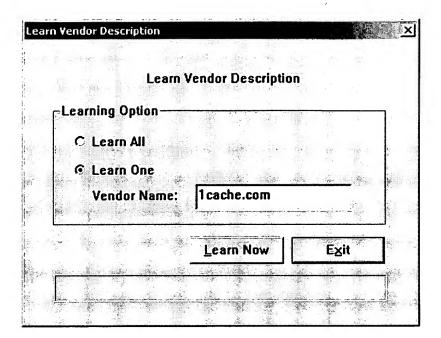


FIGURE 22



Applicants: Victor Hsieh

Sheet 22 of 36

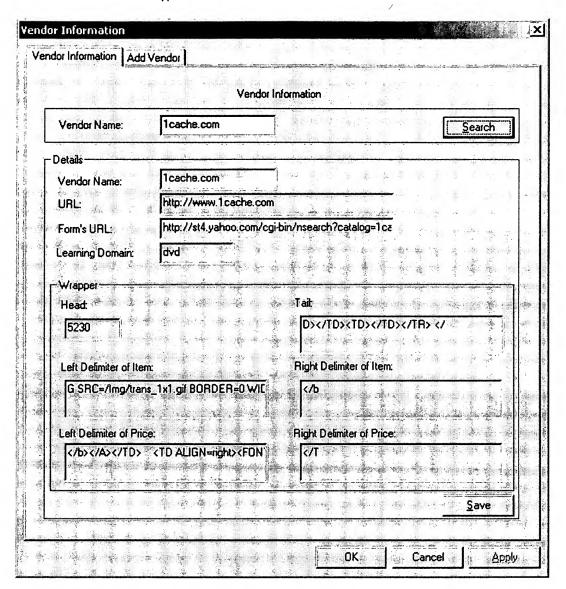


FIGURE 23



new U.S. Appln. based on Appln. Nos. 60/236,574 & 60/299,360 for Online Intelligent Information... Express Mail Label No. EL864389877US

Express Mail Label No. EL8643898770 Attorney Docket No.: 2102680-990101 Gray Cary et al. – GTS/415-836-2500

Applicants: Victor Hsieh Sheet 23 of 36





FIGURE 24

```
procedure exec_{HLRT} (wrapper (h, l<sub>i</sub>, r<sub>i</sub>, l<sub>p</sub>, r<sub>p</sub>, t), page P)

m \leftarrow 0

scan in P to line h

while the next occurrence of l_i in P occurs before the next occurrence of t

m \leftarrow m + 1

scan in P to the next occurrence of l_i in P; save position as b_{m,i}

scan in P the next occurrence of r_i; save position as e_{m,i}

scan in P to the next occurrence of l_p in P; save position as b_{m,p}

scan in P to next occurrence of r_p; save position as e_{m,p}

return label \{\ldots, << b_{mi}, e_{mi}>, < b_{mp}, e_{mp}>> \}
```

FIGURE 25

Procedure learnHLRT (examples ε)

- 1. Generate the candidate sets $Cands_r(I, \varepsilon)$, $Cands_l(I, \varepsilon)$, $Cands_l(p, \varepsilon)$, $Cands_l(p, \varepsilon)$.
- 2. Enumerate the cross product of these candidate sets; each element $W = (h, l_i, r_i, l_p, r_p, t)$, of this cross product is a wrapper. Halt if W is satisfactory i.e., execLR(W, P_n) = Ln, for every $(P_n, Ln) \in \epsilon$.

FIGURE 26



Applicants: Victor Hsieh

Sheet 24 of 36

```
* The main problem
 */
    procedure learnHLRT (examples ε) for each 1 <k <K
      for each u \in cands_i(p, \varepsilon): if valid_i(u, p, \varepsilon) then l_p \leftarrow u and terminate this loop
      for each u \in cands_r(i, \varepsilon): if valid_r(u, i, \varepsilon) then r_i \leftarrow and terminate this loop
      for each u \in \text{cands}_r(p, \varepsilon): if \text{valid}_r(u, p, \varepsilon) then r_p \leftarrow u and terminate this loop
      for each u_{ii} \in cands_i(i, \varepsilon)
      for each u_h \in cands_h(\varepsilon)
         for each u_t \in cands_t(\varepsilon)
             if valid_{ih,t}(u_{li}, u_{h}, u_{t}, \epsilon) then
                 I_i \leftarrow u_{li}, h \leftarrow u_{h}, t \leftarrow u_{t}, and terminate these three loops
    return HLRT wrapper (h, l<sub>i</sub>, r<sub>i</sub>, l<sub>p</sub>, r<sub>p</sub>, t)
 *Generate a set of candidates for left delimiter of price attribute
procedure cands<sub>i</sub>(attribute price, examples ε)
 return the set of all suffixes of the shortest string in neighbors (price, \varepsilon)
 *Generate a set of candidates for right delimiter of attribute a. Here a could be item or price
procedure cands<sub>r</sub>(attribute a, examples ε)
      return the set of all prefixes of the shortest string in neighbors, (a, \varepsilon)
 * Generate a set of candidates for a page's head
procedure cands<sub>h</sub> (examples \varepsilon)
      return the set of all substrings of the shortest string in heads(ε)
 *Generate a set of candidates for a page's tail
procedure cands, (example \varepsilon)
      return the set of all substrings of the shortest string in tails(ε)
 *Determine whether a particular candidate for the left delimiter of price is valid
 *This procedure applies constraints C3
   procedure valid(candidate u, attribute price, examples ε)
             for each s∈ neighbors₁(price, ε): if u is not a proper suffix of s then return FALSE
```

Figure 27A





```
for each s \in tails(\varepsilon): If u is a sub string of s then return FALSE
        return TRUE
 *Determine whether a particular candidate for the right delimiter of item or price is
 *This procedure applies constraints C1 and C2
   procedure valid<sub>ε</sub>(candidate u, attribute a, examples ε)
for each s \in attribs(a, \varepsilon): if u is a sub string of s then return FALSE
for each s \in \text{neighbors}(a, \varepsilon): if u is not a prefix of s then return FALSE
   return TRUE
 *Determine whether a particular combination of candidate uh, ut, and uli for h, t, and
 *li respectively are satisfactory.
procedure valid<sub>ii,h,t</sub>(candidates u_{ii}, u_{h}, u_{t}, example \varepsilon)
   for each s \in heads(\varepsilon)
        if is not a sub string of s then return FALSE
     if U<sub>11</sub> is not a proper suffix of scan (s, Uh) then return FALSE
            if Ut occurs before U11 in scan (s, Uh), then return FALSE
   for each s \in tails (\varepsilon)
     if Ut 1S not a sub string of s then return FALSE
        if Uli occurs before Ut in s then return FALSE
   for each s \in seps(\varepsilon)
        if Uli is not a proper suffix of s then return FALSE
            if Ut occurs before Uli in s then return FALSE
   return TRUE
 *Return a set of containing all values of a attribute, either item(i) or price(p) in each
 *example
*/.
   procedure attribs(attribute a, examples ε)
            \text{return } U_{(Pn,Ln)} \ \in \epsilon(Pn[b_{m,a},\ e_{m,a}] \ I < \ldots, \ (b_{m,k}, \, e_{m,k}), \, \ldots \, > \, \in \, Ln \}
 *Return all strings to the left of an attribute, whether these strings are in the heads
 * or the bodies of the pages
*/
   procedure neighbors<sub>i</sub>(attribute a, examples \varepsilon)
            if a=i then return seps (i, \varepsilon) U heads (\varepsilon) else return seps (a, \varepsilon)
```

FIGURE 27B



NEXTGEN.ASP BROWSER PROJECT.ASP MSWCRUM.DLL INTERNET **WEBCLASS ACTIVE SERVER INFORMATION RUNTIME PAGES SERVER** APPLICATION OBJECT **BROWSERTYPE OBJECT REQUEST OBJECT RESPOND OBJECT SERVER OBJECT SESSION OBJECT WEB SERVER**

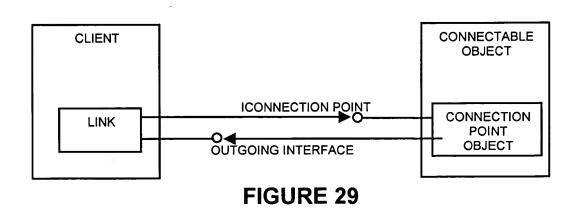
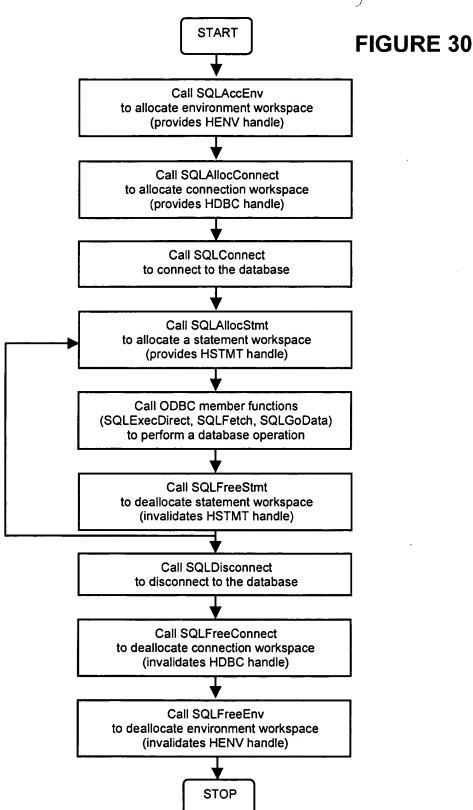


FIGURE 28

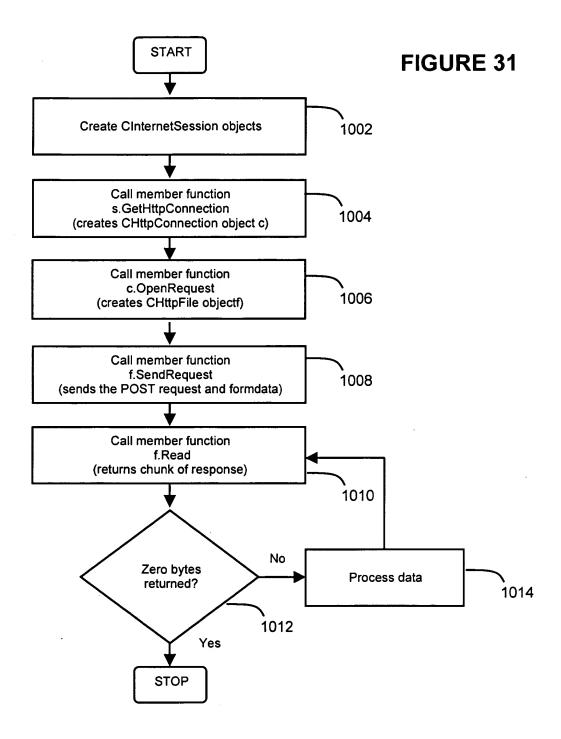
Applicants: Victor Hsieh Sheet 27 of 36





Applicants: Victor Hsieh

Sheet 28 of 36



	NAVIGATION BUTTONS TO OTHER PAGES IN WEBSITE	
WEBSITE LOGO	C All Ontological Domains	Choose a Channel ains
		Advanced Search
QUICK SEARCH		
Enter a "MULTILIGUAL CHARACTER" keyword and click GO!	CTER" keyword and click GO!	
Keyword: GOJ		
MEMBER SIGN IN	WEBSITE INTRODUCTORY TEXT AND IMAGES	Reverse Advertising Sites
Login Reset		
Registered Vendor Links Abcd.com cdef.com 1234z.com 23ezidcom	(P)Feedback	
pkkw.com er5lcin.com	Submit	

FIGURE 32

new U.S. Appln. based on Appln. Mos. 60/236,574 & 60/299,360 for Online Intelligent Information... Express Mail Label No. EL864389877US
Attomey Docket No.: 2102680-990101
Gray Cary et al. – GTS/415-836-2500
Applicants: Victor Hsieh Sheet 29 of 36

	NAVIGATION BU	NAVIGATION BUTTONS TO OTHER PAGES IN WEBSITE	GES IN WEBSITE	
				Choose a Channel
WEBSITE	LOGO		ALL G2B PROCUREMENT DOMAINS $oldsymbol{\nabla}$	MENT DOMAINS
				Advanced Search
DOMAIN A	DOMAIN B	DOMAIN C	DOMAIN D	OTHER DOMAINS
QUICK SEARCH				
Enter a "MULTILIGUAI	AL CHARACTER" keyword and click GO!	and click GO!		
Keyword:	Company A Infomation about G2B Company A	ompany A		
SEARCH TIPS	Company G Information about G2B Company G	Company G		Reverse Advertising Sites
Registered Vendor Links Abcd.com cdef.com 1234z.com 23ezidcom	Links Om L.com			
pkkw.com er5lcin	Submit.			

FIGURE 33

new U.S. Appln. based on Appln. Mos. 60/236,574 & 60/299,360 for Online Intelligent Information... Express Mail Label No. EL86438987105 Attomey Docket No.: 2102680-990101 Gray Cary et al. – GTS/415-836-2500 Applicants: Victor Hsieh Sheet 30 of 36

	NAVIGATIC	N BUTTONS T	о отнев РАС	NAVIGATION BUTTONS TO OTHER PAGES IN WEBSITE	
WEBSITE	0907			Choose a Cha	Choose a Channel
				Adv	
DOMAIN A	DOMAIN B	00	DOMAIN C	DOMAIN D	OTHER DOMAINS
ADVANCED SEARCH	CH AGENTS ARE ON!				
Enter a "MULTILIGL	Enter a "MULTILIGUAL CHARACTER" keyword and click GO!	vord and click G	iOi		
	Select All	Multilin	Multilingual Government Networks	nt Networks	
	Public Body	Speed	Air/Shipping/G	Air/Shipping/Ground/Free Overnight Del.	Ease of Navigation
i i	☐G2B Company G	በሰሰሰ		ſî	EEEE
WEAKCH TOXACH TO					
	Time Out				
	Reserved Price Range	From \$	To \$	(F)	
	Manufacturer or OEM				
	Keyword:				

FIGURE 34

NAVIGATION BUTTONS TO OTHER PAGES IN WEBSITE	NS TO OTHER P	AGES IN WEBSITE	
WEBSITE LOGO		All Ontological Domains	Choose a Channel nains
1			Advanced Search
DOMAIN A DOMAIN B	DOMAIN C	DOMAIND	OTHER DOMAINS
QUICK SEARCH			
Enter a "MULTILIGUAL CHARACTER" keyword and click GOI	click GO!		
SEARCH TIPS * SPARCH TIPS * SPARCH TIPS * SUPPLIES * * * * * * * * * * * * * * * * * *	5	More Business to Business MRO/Raw Materials Office Furnishings Office Supplies Real Estate Restaurant/Hospitality Services	Reverse Advertising Sites
o Li			
pkkw.com erbicin.com Submitt			

new U.S. Appln. based on Appln. Wos. 60/236,574 & 60/299,360 for Online Intelligent Information... Express Mail Label No. EL864389877US Attomey Docket Wo.: 2102680-990101 Gray Cary et al. – GTS/415-836-2500 Applicants: Victor Hsieh Sheet 32 of 36

FIGURE 35

	NAVIGAŢI	ON BUTTONS TO	О ОТНЕЙ РАС	NAVIGATION BUTTONS TO OTHER PAGES IN WEBSITE	
				Choo	Choose a Channel
WEBSITE	L060			All Ontological Domains	D
				Adva	Advanced Search
DOMAIN A	DOMAIN B	IOO	DOMAIN C	DOMAIN D	OTHER DOMAINS
ADVANCED SEAR(ADVANCED SEARCH AGENTS ARE ON!				
Enter a "MULTILIG	Enter a "MULTILIGUAL CHARACTER" keyword and click GO!	word and click G	ō		
	Select All	Multilin	Multilingual B2B Suppliers	liers	
	Online Stores	Speed	Air/Shipping/	Air/Shipping/Ground/Free Overnight Del.	Ease of Navigation
	☐B2B Company 1	nnnn		î	EEEE
SEARCH LIPS	☐ B2B Company 2	በበበበበ		↑	EEEEE
* +	☐ B2B Company 3	በበበበ		î	EEEE
	☐ B2B Company 4	ກກກກ		ſſ	EEEE
	Time Out				
	Price Range	From \$		₩	
	Manufacturer				
	Keyword:			(00)	

FIGURE 36

new U.S. Appln. based on Appln. Mos. 60/236,574 & 60/299,360 for Online Intelligent Information... Express Mail Label No. EL864389877US
Attorney Docket Mo.: 2102680-990101
Gray Cary et al. – GTS/415-836-2500
Applicants: Victor Hsieh Sheet 33 of 36

	NAVIGAŢ	NAVIGATION BUTTONS TO OTHER PAGES IN WEBSITE	ER PAGES IN M	/EBSITE
				Choose a Channel
WEBSITE LC	090			All Domains
				Advanced Search
DOMAIN A	DOMAIN B	DOMAINC		DOMAIN D OTHER DOMAINS
QUICK SEARCH				
Enter a "MULTILIGUAL (CHARACTER" ke	L CHARACTER" keyword and click GO!		
Keyword:		Hot Sellers Domain A	omain A	
SEARCH TIPS	Image Item A	Description Item A	Image Item B	Description Item B
* * *	Image Item 1	Description Item 1	Image Item G	Description Item G
	Image Item A	Description Item A	Image Item K	Description Item K
Registered Vendor Links Abcd.com cdef.com	s Feedback	edback		
1234z.com 23ezidcom pkkw.com er5lcin.com	m Submit			

FIGURE 37

	NAVIGAŢ	ON BUTTONS T	NAVIGATION BUTTONS TO OTHER PAGES IN WEBSITE	
			3	Choose a Channel
WEBSITE	LOG0		All Domains	⊳
			1	Advanced Search
DOMAIN A	DOMAIN B	OQ	DOMAIN C DOMAIN D	OTHER DOMAINS
ADVANCED SEARCH	CH AGENTS ARE ON!			
Enter a "MULTILIGI	Enter a "MULTILIGUAL CHARACTER" keyword and click GO!	word and click G	iO	
	Select All		English Vendors	
	Online Stores	Speed	Air/Shipping/Ground/Free Overnight Del.	el. Ease of Navigation
	☐ Store 1AAC	UUU	ſì	EEEE
SEARCH TIPS	☐ Store 1443A	ມມມມ	î	EE
* •	☐ Store 1443A	ՈՈՈՈ	Î	33
	Select All		Chinese Vendors	
	Ctory CEA3	00000		U U
	C Signe CoAs		↑	בבבר
	☐ Store DRGWS4	nnnnn	Î	EEEE
,	Time Out			
	Price Range	From \$	To \$	
	Manufacturer			
	Keyword:			

FIGURE 38

	NAVIGATION B	NAVIGATION BUTTONS TO OTHER PAGES IN WEBSITE	AGES IN WEBSITE	
			0	Choose a Channel
WEBSITE	LOGO		ALL DOMAINS	٥
				Advanced Search
DOMAIN A	DOMAIN B	DOMAIN C	DOMAIN D	OTHER DOMAINS
QUICK SEARCH				
Enter a "MULTILIGUAI	AL CHARACTER" keyword and click GO!	1 and click GO!		
Keyword:	Search Results			
OS GIF HOGVES	Products Your search for XX retu	Products Your search for XX returned 5 Domain C items.	:	
	Model / Features	res	Store	Price
× •	AMS7 XX Widget [Features]		Store 1443A	\$399.95 Buy now
	AMS2 XX Widget [Features]		Store 1443A	\$329.95 Buy now
	AMS7 XX Widget [Features]		Store DRGWS4	\$459.95 Buy now
	AMS7 XX Widget [Features]		Store 4U\$U\$	\$499.95 Buy now
	ZRTQ XX Widget [Features]		Store OOHE	\$529.95 Buy now

FIGURE 39

new U.S. Appln. based on Appln. Nos. 60/236,574
& 60/299,360 for Online Intelligent Information...
Express Mail Label No.: E186438987105
Attorney Docket No.: 2102680-990101
Oray Cary et al. – GTS/415-836-2500
Applicants: Victor Haieh Sheet 36 of 36